DOCUMENT RESUME

ED 256 780 TM 850 220

AUTHOR Flanery, Randall C.
TITLE Psychometric Issues: Is Validity Relevant to Structured Interviews?
PUB DATE Aug 84
NOTE 23p.; Paper presented at the Annual Meeting of the American Psychological Association (Toronto, Canada, August 24-28, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adolescents; Children; Evaluation Criteria; *Interviews; Predictive Validity; *Psychological Evaluation; Psychometrics; Psychopathology; *Test Validity
IDENTIFIERS Child Assessment Schedule; Diagnostic Interview for Children and Adolescents; Interview Schedule for Children

ABSTRACT

Increasing attention is being given to the psychometric adequacy of recently published structured child interviews. Three published child interviews, The Child Assessment Schedule (CAS), The Diagnostic Interview for Children and Adolescents (DICA), and The Interview Schedule for Children (ISC) were evaluated in terms of the available evidence of their validity. Four types of validity were examined: content validity; criterion-related validities; concurrent and predictive validity; and construct validity. A major obstacle to validity investigations has been the absence of a widely agreed upon criterion which is truly more valid than the interview itself. This issue has been dealt with by studying known groups, by comparing the degree of association between interview scores and other measures of psychopathology, and agreement between child and parent on interview. All three interviews were judged to show reasonable evidence of content validity, although there are more questions regarding the DICA. The most common validity investigation has been of concurrent validity. The few studies of the ISC and CAS have found adequate validity. Studies of the DICA are of poor quality and have produced disappointing results. The predictive or construct validities of these interviews are virtually uninvestigated. The preliminary studies are encouraging but are of insufficient quantity to definitively judge their validity. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED256780

Psychometric Issues:

Is Validity Relevant to Structured Interviews?

Randall C. Flanery, Ph.D.

Department of Pediatrics

University of Wisconsin Hospital

Presented at the Annual Meeting of the American Psychological Association, August 24-28, 1984, Toronto, Canada.

TM 850 220

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.C. Flanery

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Abstract

Increasing attention is being given to the psychometric adequacy of recently published structured child interviews. Three published child interviews, The Child Assessment Schedule (CAS), The Diagnostic Interview for Children and Adolescences (DICA), and The Interview Schedule for Children (ISC) were evaluated in terms of the available evidence of their validity. Four types of validity are examined: content validity; criterion-related validities; concurrent and predictive validity; and construct validity. A major obstacle to validity investigations has been the absence of a widely agreed upon criterion which is truly more valid than the interview itself. This issue has been dealt with by studying known groups, by comparing the degree of association between interview scores and other measures of psychopathology, and agreement between child and parent on interview. The three interviews were all judged to show reasonable evidence of content validity, although there are more questions regarding the DICA. The most common validity investigation has been of concurrent validity. The few studies of the ISC and CAS have found adequate validity. Studies of the DICA are of poor quality and have produced disappointing results. The predictive or construct validities of these interviews are virtually uninvestigated. The preliminary studies are encouraging but are of insufficient quantity to definitively judge their validity.

BEST COPY

The advent of structured interviews designed to assess the psychological status of children (e.g., Herjanic, Herjanic, Brown & Wheatt, 1975; Hodges, Kline, Stern, Cytryn, & McKnew, 1982; Kovacs, 1978) has stimulated interest in the psychometric qualities of these instruments. Interest has focussed primarily on the reliability of these measures although increasing attention is being devoted to validity concerns. While consumers of structured child interviews must applaud these efforts, far greater information is both desired and required. The APA Standards on Tests and Measures states clearly that "It is intended that these standards apply to any assessment procedure, assessment device, or assessment aid; that is, to any systematic basis for making inferences about characteristics of people (p. 2)." Interviews are systematized measures for obtaining data necessary to make judgements about diagnosis, treatment, etc. In other words, structured child interviews are exactly the kinds of measures the authors of the Standards had in mind. Developers of structured child interviews have a primary ethical responsibility to marshall evidence of the validity as well as the reliability of their instruments. The plan of this article is to examine ways in which psychometric concepts of validity may be appropriately applied to structured child interviews and then evaluate three published child interviews in light of these notions of validity. Validity has been chosen for emphasis since the reliability of the measures has received greater attention.

While it may be argued that standards of validity which have evolved from traditional psychometric theory may be applied to child interviews, that is not to say that structured interviews are tests in the conventional sense. Structured interviews do not assess unitary, trait-like constructs. Indeed the opposite is true. Most diagnostic interviews

BEST COPY

are designed to assess a broad range of content in as parsimonious a fashion as possible. Items are chosen to be maximally unique, that is, assessing non-overlapping material. Interview items which share common content and are most likely to be inter-correlated would be eliminated, whereas items in a traditional test are specifically chosen because they are moderately correlated, i.e., measure related content.

Furthermore, interview items are more likely to be selected for theoretical or clinical reasons than on empirical grounds. Thus, many questions are included because they contribute to making a diagnosis or because the test constructor believes it to be important. By contrast, a psychometrician might exclude items based on an unsatisfactory response distribution or because it correlates poorly with the total scale score, regardless of its importance on other grounds.

Interview items are likely to be very heterogeneous, in contrast to the relative homogeneity of test items. Interview items vary in their degree of concreteness, ranging from "Do you wet the bed?" to "How much do you feel lonely?" Another source of heterogeneity is the variety of responses a child can give which must then be coded by the interviewer into "present" or "absent." Even with explicit guidelines, one would expect substantial variability in clinical judgements of the presence of symptoms.

The ways in which interview questions differ from test items make interview data less psychometrically sound, thus increasing the difficulty of establishing the validity of an interview. The differences enumerated above are sources of variance which, because they are treated as error variance, could easily obscure real relationships. Because the purposes of an interview are sometimes at odds with methods of good

BEST COPY

test construction, structured interviews often look like poorly constructed tests to rigorous psychometricians. The dilemma facing developers of interviews is illustrated by the problem of trying to reduce sources of heterogeneity. A test can be improved psychometrically by increasing the number of items or by eliminating statistically poor items. It is unlikely that an item would be added to an interview unless it solicited new information, nor would a question which is relevant to some diagnostic or treatment consideration be removed on statistical grounds.

One effect of the technical difficulties attendant upon constructing a good interview is that it reinforces the widely held belief that notions of test validity do not apply to structured interviews (although this belief has been formally challenged of late, e.g., Kazdin & Petti, 1982). The belief is further underscored by the fact that very often the criterion in the validation of other measures, especially measures of psychopathology, is provided by an interview. For example, measures of depression are demonstrated to be valid by showing that an individual diagnosed as depressed via an interview scores high on the depression measure whereas someone who is diagnosed as not depressed scores low. In other words, the interview is construed as the more valid measures. Using psychometric terminology, structured interview data are by definition the criterion to be predicted by other, non-interview measures. By this reasoning, questions of validity apply not so much to the interview as to the predictors of interview data. However to quote once again from the APA Standards (p. 27), "...the logic of criterion-related validity assumes that the criterion possesses validity. All too often tests are validated against any available criterion with no corresponding investigation of the criterion itself. The merit of a criterion-validity study depends on the appropriateness and quality

BEST COPY

of the criterion measure chosen." To the extent that child interviews are to be employed as a criterion against which other instruments are to be measured, questions of validity become all the more important. As if the technical challenges were not enough; it turns out that one must not only overcome the many technical challenges such as sources of heterogeneity, the interview developer must also seize upon criterion more valid than an interview.

Structured Child Interviews

Three structured child interviews were evaluated: the Child Assessment Schedule (CAS; Hodges et al., 1982), the Interview Schedule for Children (ISC; Kovacs et al., 1984), and the DICA (Herjanic et al., 1975). Detailed descriptions of the interviews have been published only recently, so it is not surprising that validity data is relatively scanty. While more rigorous validity studies have yet to appear but are presumably underway, enough data have been published to draw some preliminary conclusions.

Child Assessment Schedule (CAS). The CAS (Hodges et al., 1982) consists of two parts. The first part consists of 75 questions about a variety of topics including school, friends, activities, fears, mood, somatic complaints, expression of anger and thought disorder symptomatology. Examples of responses are provided to aid classification by the interviewer. Responses are coded as "No" or "false", "Yes" or "true", Ambiguous, No Response or Not Applicable. The second part has 53 behavioral observation items to be completed after the interview is done. These include grooming, insight, motor coordination, activity level, etc. Scores derived from the CAS are total number of symptoms, number of symptoms in 11 content areas and 9 DSM-III diagnostic symptom complexes. It is

BEST COPY

designed to be used with children aged 7 to 12 years.

Diagnostic Interview for Children and Adolescents (DICA). The DICA (Herjanic et al., 1975) consists of 207 questions which generally may be answered "yes" or "no". Factual information such as age constitute 20 questions. Behavior in several contexts such as at home or school are the focus of 79 questions. Psychiatric symptoms are assessed in 102 questions. Six factors summarize the mental status of the child. A total symptom score and symptom scores for six areas, relationship problems, school behavior, school learning, neurotic symptoms, somatic symptoms and psychotic symptoms are calculated. Diagnoses are also obtained, which differ markedly from DSM-III diagnoses. Which symptoms and how they are combined to make a particular diagnosis are not stated explicitly. The interview may be used with children aged 6 to 16 years.

Interview Schedule for Children-Form C (ISC). The ISC (Kovacs et al., 1978) contains 71 items which assess psychopathology, mental status, developmental milestones, behavioral observations and clinician impressions. DSM-III diagnoses may be made. Most items are rated on a 0 to 8 scale of severity, with each point defined. A few items are on a 0 to 3 scale. Inquiries of the child are provided and guidelines for conducting the interview. Children aged 8 to 17 years can be interviewed with the ISC.

Evidence of Validity

Questions of validity may be subsumed by two distinct but related questions: 1) How well does the instrument measure the skill, trait, or domain that it is alleged to measure, and 2) How well does the instrument predict performance in some other domain or on some other indicator of a skill or trait? Many terms are used and numerous procedures described in discussions of validity. The basic inferential decisions for summarizing

BEST COPY

the adequacy of a measure, regardless of methods, are encompassed by four kinds of validity: content validity; criterion-related validities, namely concurrent validity and predictive validity; and construct validity.

The current status of structured child interviews as assessment devices can be evaluated by examining evidence of these four types of validity. No single type of validity is usually sufficient to conclude that an interview is adequate. Rather, comprehensive analysis of the relationship among different types of validity is necessary, with careful consideration of the specific purposes and contexts in which the interview is to be used. This is generally a lengthy process and requires more than a single study.

The adequacy of the three interviews (CAS, DICA, ISC) was evaluated in light of evidence of the four traditional types of validity. More generally, the questions were: 1) How well does the interview assess the psychological status of the child, and 2) How well do responses to the interview correspond to behavior in other contexts and to other indices of psychological dysfunction or well-being? Evidence supporting the validities of the three interviews are summarized in Table 1.

INSERT TABLE 1 HERE

Content Validity

A child interview may be considered to be content valid if the specific items or questions fully represent the universe of behaviors, feelings, etc. of interest. According to Cronbach (1971), content validity requires 1) an unambiguous definition of the relevant universe of content, 2) the measure's items adequately sample or represent the specified universe,

BEST COPY

and 3) the method for classifying responses and combining them into scores can be stated clearly. The criteria emphasize that care should be given equally to the items and to the responses. This is especially important for structured interviews with children when one often wonders whether the child has understood the question as was intended. Even when the item seems to have been understood, it is often difficult to interpret the response unambiguously. Furthermore, considerable latitude is often given to the interviewer to repeat or paraphrase the item in order to more clearly ascertain the meaning of the child's response. As the true meaning of the child's statement is pursued, the clarity of the item and the method for classifying the response begin to dissipate. Unfortunately, there are no post hoc statistical procedures for estimating content validity. It is a concern which is most salient during the construction of a measure. Content validity is assessed primarily by evaluating the procedures used to select items and by comparing the final product to the explicitly stated definition of the universe of interest.

The ISC is especially noteworthy for describing in detail the manner in which it was developed. A clear description of guidelines, sources, scoring procedures, etc. are critical for evaluating content validity. The considerations which guided changes in format, inclusion and exclusion of items and the rationale for the system of coding child responses are documented (Kovacs, 1973; 1983). The construction of the CAS is less well described but the origin of items is stated. Like the ISC, the CAS includes precise guidelines and examples for unambiguously classifying child responses. For example, the presence of a symptom is defined concretely, and responses may be scored as Absent (No), Present (Yes), Ambiguous, Not Asked, or No Response. Alternatively, the ISC typically provides

BEST COPY

a 0 to 9 rating scale with each point fully defined. By contrast, the construction of the DICA is virtually unknown. It is not clear where the items actually came from, on what basis, if ever, items were included or excluded, although an examination of items suggests that most relevant areas of child behavior are assessed. Unlike the ISC and CAS, the DICA does not include guidelines for classifying a particular behavior as present nor are principles for conducting the interview ever stated. This can be a particular problem when the severity of a problem behavior is crucial to a judgement. For example, many children apparently have fears (Achenbach & Edelbrock, 1981) but which do not necessarily constitute a problem. The DICA would be unlikely to differentiate among these types of child problems.

A DSM-III diagnosis is often an expected outcome of administering a child interview. If so, data necessary to make a diagnosis must be elicited by the interview. The CAS and ISC were explicitly designed to obtain data for DSM-III diagnoses. The DICA also provides diagnostic data (Herjanic et al., 1982) but they are not DSM-III diagnoses. To the extent that interview items correspond to diagnostic criteria, content validity can be presumed to exist. In this context, concerns about items are primarily criticisms of diagnostic criteria rather than evidence of a lack of content validity for the interview. Depending upon one's judgement of the validity of the diagnostic criteria, all three interviews are probably content valid, although a preference would be given to DSM-III criteria by virtue of its wider usage.

In sum, all three interviews probably are content valid. This conclusion can be made about the CAS and the ISC, with relatively little qualification since test construction was described fairly comprehensively

BEST COPY

and test administration and scoring procedures are very well described. The DICA is poorly documented and guidelines for usage seem especially lacking, so questions about content validity remain. This is ironic because of the three interviews, the most articles have been published about the DICA.

All three interviews focus primarily on psychopathology, which may pose a threat to content validity. A critic might charge that an exclusive emphasis on symptoms neglects the importance of competencies to the psychological well-being of a child. In part this is a definitional issue. A proponent of these interviews could argue that the chosen universe of interest was defined so as not to include child competencies. In the absence of an agreed upon definition of the universe of interest, such questions are unresolvable.

Criterion-Related Validities

Criterion-related validities are examined when performance on a structured child interview is shown to be associated with standing on some other variable, the criterion, which is assumed to be more valid. If the inference is about an individual's status on the criterion in the future, then predictive validity is an issue; if current status on the criterion is of interest, then concurrent validity is being studied. Criterion-related validity is demonstrated by showing that individuals' test scores are highly correlated with the same individuals' standings on some other variable, the criterion. A major hinderance in studies of this kind is that an interview is often seen as the more valid criterion against which to compare other measures. As such it becomes difficult to identify a criterion against which to compare a structured child interview.

BEST COPY

A common procedure for coping with the valid criterion problem has been to attempt to use the interview to differentiate "known groups" which have been defined previously on some other basis. The assumption is that the method of differentiation is indeed valid, about which there may be some question in this case. One definition of "known groups" has relied upon patient status rather than diagnosis. Herjanic and Campbell (1977) recruited 50 children from a pediatric clinic and compared them on the DICA to a sample of 50 children evaluated at a psychiatric clinic. It was concluded that the DICA could distinguish the more psychologically disturbed children of the psychiatric sample from the less disturbed, pediatric sample. Forty-one psychiatric children had more positive symptoms on the DICA than their matched pediatric child. Means in six symptom areas were higher for the psychiatric sample than the pediatric sample although statistics were not reported. While the study reports useful validity data, it used a number of questionable statistical procedures. Evidence of concurrent validity is suggested, but the demonstration is less than rigorous.

The CAS has discriminated among psychiatric inpatient, outpatient, and normal children (Hodges et al., 1982). The psychiatric sample (inpatient and outpatient) were significantly different from normal controls or total number of symptoms endorsed, 7 of 11 content area scores and 6 of 9 symptom complex scores (which correspond to DSM-III diagnoses). Furthermore, inpatients differed from outpatients, who in turn differed from controls on total number of symptoms, one content area (expressions of anger), and 4 symptom complex scores. This was better than the performance of the Child Behavior Checklist (CBC, Achenbach, 1978), which could distinguish psychiatric patients and controls, but could not differentiate

BEST COPY

inpatients and outpatients. The superior discriminating ability of the CAS compared to the CBC is a powerful argument for the necessity of structured interview data. This is a more powerful test of the concurrent validity of the CAS, requiring more specific discrimination than simply distinguishing non-disturbed and disturbed children. The ISC apparently has not yet been evaluated using known groups.

The demonstrations of concurrent validity of the type described above relied on a discrimination of the very general dimension of psychopathology. While serving as a necessary indicator of validity, it is of greater import to know if an interview can serve as a precise diagnostic assessment rather than simply as an indicator of the severity of psychological impairment. For example, can individuals be assigned unambiguously to discrete diagnoses? The problem of a diagnostic criterion is, of course, a major issue. Such a demanding investigation has not been conducted.

Another method for evaluating concurrent validity is to obtain correlations between interview scores and other measures. The CAS is the only interview of the three to have been evaluated in this way (Hodges et al., 1982). CAS symptom complex scores were found to correlate with CBC scale scores which assessed comparable content. For example, the CAS depression complex score correlated ($r=.53$, $p < .001$) with the Child Depression Inventory (CDI, Kovacs & Beck, 1977).

An alternative method is to compare child data with parental responses to the same interview questions. High agreement between child and parental report would be expected since the content and method are identical but the informant is different. This kind of data is often reported as reliability data. The distinction between reliability evidence and validity

BEST COPY

evidence is often arbitrary. Such data can be seen as similar measurements of the same thing (i.e., reliability) or as similar measurements of different variables (i.e., validity). The latter understanding is more appropriate because what is actually being compared is the child's self-perceptions to the parent's perception of the child. In this light, parental report is a different variable rather than the measurement of the same variable.

The DICA and the ISC have reported highly variable agreement between parent and child reports of structured interview content. Kovacs (1983) reported correlations between child and parent report which ranged between .02 for child's pessimistic thinking and .95 for truancy. The more external and objective the target symptom, the greater was parent-child agreement. The inescapable conclusion is that child and parent reports are not interchangeable (i.e., similar measures of the same thing; also called reliability), especially for more internal subjective events such as mood and cognitions. Similar results have been reported for the DICA (Herjanic et al., 1975; Herjanic & Reich, 1982). Items assessing objective, concrete events had high agreement whereas affective symptoms showed virtually no agreement. An examination of DICA (non-DSM-III) diagnoses based on parent and child data showed similar lack of agreement (Reich et al., 1982). The kappas, which are an estimate of concordance with a correction for chance agreement, obtained in this study were fairly low. Only two exceeded .50, encopresis and antisocial personality, diagnoses relying upon identification of concrete events. Other kappas were below .40. This is very low agreement for the most part. The DSM-III field trials considered a kappa of .7 to be high agreement (American Psychiatric Association, 1980). Most kappas for specific DSM-III diagnoses were above .44. The concurrent validity of the DICA has not been definitively demonstrated.

BEST COPY

The results described above present a mixed picture as to the concurrent validity of these measures. Satisfactory agreement was found for items which were most objective but agreement was disappointing for more internal content. The latter are the areas of most interest and one important reason for using structured child interviews at all. It could be concluded that these data indicate a lack of concurrent validity, which is predicated on the notion that the criterion, parental report, is more valid than the predictor, the child interview. It clearly is not. A more definitive study requires a better criterion, which has yet to be identified.

The studies conducted to date have been of concurrent validity. Evidence of predictive validity, i.e., attempting to predict future psychological status, may be found only for the ISC. In a recent, initial report of a longitudinal study of childhood depressive disorders (Kovacs, Feinberg, Crouse-Novak, Paulauskas, & Finkelstein, 1984), the ISC identified children with a major depressive disorder (MDD), dysthymic disorder (DD), or adjustment disorder with depressed mood (ADDM). The results indicated differences among the three groups on age of onset, mean episode length and recovery rate. In an impressive display of diagnostic discrimination, meaningful and different patterns of responses were documented by the ISC for each diagnostic groups. The sensitivity of the ISC to temporal variation in symptoms manifested by children with different diagnoses may be taken as evidence of predictive validity.

Construct Validity

Conclusions about the construct validity of a measure are based upon the accumulated evidence of many studies and the corresponding development of a theory to account for results. The purpose of studies examining structured interviews have been more empirical than theoretical. The

BEST COPY

construction of the interviews have been guided by pragmatic concerns rather than by explicitly theoretical considerations. It is certainly true that the kind of comprehensive theory of child behavior necessary for evaluating construct validity has not been articulated. The beginnings of some hypotheses may be beginning to appear. The intriguing discrepancies in parent-child agreement and the differential time patterns for childhood depression are two areas in which an understanding of the phenomena could enhance the construct validity of the measure.

Summary

Evidence of validity of three structured child interviews (CAS, ISC, DICA) was evaluated. The CAS and ISC had ample evidence of content validity. The DICA may be content valid, although this conclusion is tentative. The fact that it is being revised to conform with DSM-III diagnoses (Reich et al., 1982) implies some inadequacy of content. Concurrent validity studies have been conducted for each interview. Some evidence of concurrent validity are available for all three measures. The available evidence of validity of the ISC and CAS is strong but the quantity of data needs to be increased. Ironically, the most extensive validity data is on the DICA, which is plagued by questionable statistical and methodological procedures. Despite the amount of data, reservations about the adequacy of the DICA persist.

As validity studies accrue, it will begin to be possible to judge the construct validity of these measures. The absence of both data and a well-articulated theory prevent useful statements about construct validity. One salutary effect of increasing attention to these measures is likely to be increased efforts to explicate the implicit theory that underlie the measures.

BEST COPY

References

Achenbach, T.M. (1978). The Child Behavior Profile: I. Boys aged 6-11. Journal of Consulting and Clinical Psychology, 46, 478-488.

Achenbach, T.M. & Edelbrock, C. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged 4-16. Monographs of the Society for Research in Child Development, No. 38, 46 (1).

Achenbach, T.M. & Edelbrock, C. (1983). Manual for the revised Children Behavior Checklist and Profile. Burlington, VT: University Associates in Psychiatry.

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, D.C.: Author.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education.

Herjanic, B. & Campbell, W. (1977). Differentiating psychiatrically disturbed children on the basis of a structured interview. Journal of Abnormal Child Psychology, 5, 127-134.

Herjanic, B., Herjanic, M., Brown, F., & Wheatt, T. (1975). Are children reliable reporters? Journal of Abnormal Child Psychology, 3, 41-48.

Herjanic, B. & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. Journal of Abnormal Child Psychology, 10, 307-324.

Hodges, K., Kline, J., Stern, L., Cytryn, L., & McKnew, D. (1982). The development of the Child Assessment Schedule for research and clinical purposes. Journal of Abnormal Child Psychology, 10, 173-189.

Hodges, K., McKnew, D., Cytryn, L., Stern, L., & Kline, J. (1982). The Child Assessment Schedule (CAS) Diagnostic Interview: A report on reliability and validity. Journal of the American Academy of Child Psychiatry, 21, 468-473.

Kazdin, A. & Petti, T. Self-report and interview measure of childhood and adolescent depression. Journal of Child Psychology and Psychiatry, 1982, 23, 437-457.

Kovacs, M. (1978a). The Interview Schedule for Children (ISC). Unpublished manuscript, West Psychiatric Institute, Pittsburgh, PA.

BEST COPY

Kovacs, M. (1978b, October). A clinical-research project on childhood depression. Paper presented at the 25th Annual Meeting of the American Academy of Child Psychiatry, San Diego, CA.

Kovacs, M. (1983, December). The Interview Schedule for Children. Unpublished manuscript, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Kovacs, M. & Beck, A.T. (1977). An empirical clinical approach towards a definition of childhood depression. In J.G. Schulterbrandt & A. Raskin (Eds.), Depression in children: Diagnosis, treatment, and conceptual models. Raven: New York.

Kovacs, M., Feinberg, T.L., Crouse-Novak, M.A., Paulauskas, S.L., & Finkelstein, R. (1984). Depressive disorders in childhood: I. A longitudinal prospective study of characteristics and recovery. Archives of General Psychiatry, 41, 229-237.

Reich, W., Herjanic, B., Welner, Z., & Gandhy, P.R. (1982). Development of a structured psychiatric interview for children: Agreement on diagnosis comparing child and parent interviews. Journal of Abnormal Child Psychology, 10, 325-336.

BEST COPY

TABLE 1

SUMMARY OF VALIDITY EVIDENCE FOR STRUCTURED CHILD INTERVIEWS

	TYPES OF VALIDITY			
	CONTENT	CONCURRENT	PREDICTIVE	CONSTRUCT
Child Assessment Schedule (CAS) (Hodges et al., 1982)	Good. Items, responses and classification of responses are well-defined.	Known Groups able to discriminate inpatient, outpatient, and controls. Parent-Child Agreement-not investigated. Association with other measures-high correlations with other measures of depression, anxiety.	Not investigated.	Insufficient data.
Diagnostic Interview for Children and Adolescents (DICA) (Herjanic et al., 1975)	Probably o.k. but some reservations especially for making diagnoses.	Known Groups-discriminated psychiatric from pediatric but few or questionable statistical procedures. Parent-Child Agreement-highly variable agreement for symptoms, diagnoses depending upon content. In general, not persuasive.	Not investigated.	Insufficient data.

BEST COPY

Table 1-Summary of Validity Evidence for Structured Child Interviews (continued)

	TYPES OF VALIDITY			
	CONTENT	CONCURRENT	PREDICTIVE	CONSTRUCT
Diagnostic Interview for Children and Adolescents (DICA) (Herjanic et al., 1975)		Association with other measures-not investigated.		
Interview Schedule for Children (ISC) (Kovacs, 1978)	Good. Items, responses and classification of responses are well-defined.	Known Groups-not investigated. Parent-Child Agreement-highly variable agreement, depending on content. Indicates non-equivalence of child and parent report. Association with other measures-not investigated.	Some preliminary data, temporal variation in symptomatology for four childhood depressive disorders.	Insufficient data.

BEST COPY